International Journal of Scientific Engineering and Technology
Volume No.3 Issue No.3, pp : 310-315
(ISSN : 2277-1581)
1 March 2014

Idea Research Based on Kernel Method in Fault Diagnosis

MohammadReza Vazifeh, Pan Hao
School Computer Science and Technology, Wuhan University of Technology, Wuhan, China
Corresponding Email: vazife66@yahoo.com

Abstract— ***it is important to reduce keeping costs and hold up unscheduled downtimes for machinery. So knowledge of what, where and how faults occur is very important. In machine rotation and machine learning Fault diagnosis and detection are important rule. In this paper offer a method based on kernel method that using in fault occur. For this reason create kernel by wavelet packet with associate rule mining and information fusion for decision rule. This kernel has best time detection and optimization misclassification. Our proposed data fusion strategies take into account that a support vector machine with multi kernel Wavelet-Entropy by finding the optimal hyper plane with maximal margin.***

Keywords— *Fault Diagnosis, Wavelet Entropy, Information Fusion, kernel method*.

I. Introduction

Numerous studies (both theoretical and empirical) have proved that are effective in achieving improved classification performance for various application problems. The failure of machinery reduces the production rate and increases the costs of production and maintenance [1]. Therefore, it is important to reduce noise and inspected event in machine learning, so knowledge of fault occur is very important.

In pattern Recognition, kernel method is a Discriminant-based classification $(g(x)|\emptyset_i)$ with linear discriminate analysis (LDA) whose suppose conditional probability is Gaussian distribution. In large data sets, best selection of kernel is important task.

In this paper we offer a new model for fault diagnosis. This research consist of 3 steps for accrue fault diagnosis based on kernel method with best position for kernel.

First step is feature extraction based on wavelet packet with associate entropy, in this step input data convert to signal model (feature map) by wavelet, and then data extract with wavelet packet tree and finally in this step select data by max entropy energy.

In second step create kernel with Mercel kernel model with *Morlet* mother wavelet on extract data for classification.

In step 3 fused data by kernel fused, in this step selecting best kernel in fusion kernel.

Our proposed data fusion strategies take into account that a support vector machine with multi kernel Wavelet-Entropy by finding the optimal hyper plane with maximal margin [2]. In the distributed schemes, the individual data sources are processed separately and modeled by using the Support Vector Machine [3]. Fault diagnosis is to detect, isolate, and assess faults and failures of engine system and its major components.

II. Material and Methodology

In pattern recognition fault is important rule. A pattern is a set of objects, processes or events which consist of both deterministic and stochastic components [4].Recognition is identification of a pattern as a member of a category that we know or we want learns (in Classification known categories and in Clustering learning categories) [5].Therefore, pattern recognition have 2 section, in pattern section make a category or class of pattern and in section of recognition make a decision about the “category” or “class” of the pattern [figure 1].

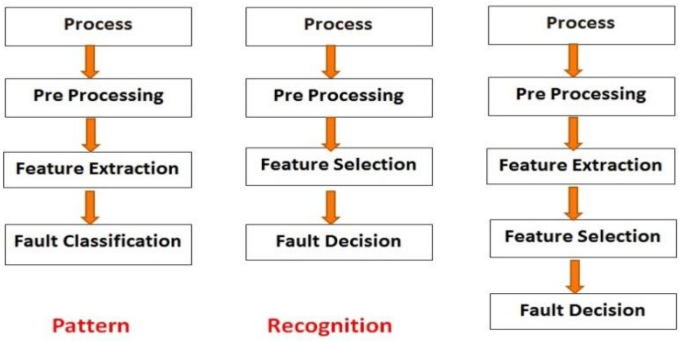

Figure 1: fault diagnosis pattern recognition

In new method for diagnosis of fault have 3 steps:

I. Feature extraction using wavelet packet with associate rule mining
II. Kernel method Classification with kernel wavelet
III. Fault decision using Information fusion by feature level fusion(kernel fusion)

I. Feature extraction based on wavelet packet transform(WPT) with associate rule mining

Feature extraction is combining attributes into a new reduced set of feature.in pattern recognition and image processing feature extraction is a reduce dimension in feature space until improve classification. Wavelet transform is powerful than other transform because wavelet transform analyze signal in both time and frequency domain.

Selection of suitable wavelet transform for given application is important, wavelet packet transform (WPT) was more suitable for understanding of the time-frequency characteristics.

Associate rule mining is a method for detection best relation between variable in large data sets.one of the quantitative measures associated with wavelet packet transform (WPT) is Entropy. Entropy can be an associate for WPT with mathematical rule. Entropy provides valuable information for analyzing non-static signals. For express the signals characteristic many various wavelet entropy presented, these entropies based on different algorithm so they have different essential meaning in application. Wavelet energy entropy is the statistical analysis of signal energy on frequency band and presents the distributing complexity of signal energy in frequency domain. Wavelet energy entropy in this paper used to obtain energy distributing information which useful in decision rule in information fusion.

II. Kernel method

The second step in fault diagnosis is classification. Support vector machine (SVM) can dodge the problems of over local minimum in the classical study method, and is applied in many classification problems successfully [6].

We assume a training set of *N* data points $\{x_k, y_k\}$ $k = 1,2,\ldots,N$, where $x_k \epsilon R^n$ is the input data, and $y_k \in R^n$ is k-th output. The SVM constructs a decision function that is showed by: $y(x) = w^T x + b$ (2-1)

In SVM for the function estimation the following optimization problem can be given [7]:

$$Min\ j_{ls}(w,b,e) = \frac{1}{2}w^T w + c\frac{1}{2}\sum_{k=1}^{N} e^2 k$$

s.t:

$$y_k = w^T x_k + b + e_k \qquad k = 1,\ldots,N \quad (2\text{-}2)$$

$$e_k, x_k \geq 0$$

Where e_k: Slack variables

c: A positive real constant

One defines the Lagrangian:

$$L(w,b,e,\alpha) = j_{ls} - \sum_{k=1}^{N} \alpha_k (w^T x_K + b + e_K - y_K) \quad (2\text{-}3)$$

With Lagrange multiplier α_k the conditions for optimality are:

$$\begin{cases} \dfrac{dl}{dw} = 0 & \rightarrow w = \sum_{k=1}^{N} \alpha_k x_k \\ \dfrac{dl}{db} = 0 & \rightarrow \sum_{k=1}^{N} \alpha_k = 0 \\ \dfrac{dl}{de_k} = 0 & \rightarrow \alpha_k \varphi e_k \\ \dfrac{dl}{d\alpha_k} = 0 & \rightarrow w^T x_K + b + e_K - y_K = 0 \end{cases} \quad (2-4)$$

Therefore we have:

$$\begin{bmatrix} I & 0 & 0 & -x \\ I & 0 & 0 & -1^T \\ I & 0 & \varphi I & -I \\ x & 0 & I & 0 \end{bmatrix} \begin{bmatrix} w \\ b \\ e_k \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ y \end{bmatrix} \quad (2\text{-}5)$$

$$\begin{bmatrix} & 0 - 1^T \\ 1 & x^T x + \varphi^{-1} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$$

With $w = \sum_N \alpha_k x_k$, $e_k = \frac{\alpha_k}{c}$ The support values α_k are proportional now to the errors at the data points.

On the basis of generalized linear critical function, we can map the input higher space into feature space by nonlinear transform to solve nonlinear problems, and evaluate optimal or generalized optimal classification plane in the feature space [6].

The kernel function can be expressed as followings:

$$K(x,x') = \emptyset(x).\emptyset(x') \qquad (2\text{-}6)$$

Then, the decision function of support vector machine can be obtained as followings:

$$f(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i t_i K(x,x') + b\right) \qquad (2\text{-}7)$$

As support vector machine can't perform the probabilistic interpretation of the decision process, relevance vector machine is applied to fault diagnosis. The likelihood of the training dataset is gained by applying the generalization linear model and logistic sigmoid function:

$$\rho(t|w) = \prod_{n=1}^{N} y(x_n,w)^{t_n}[1-y(x_n;w)]^{1-t_n} \qquad (2\text{-}8)$$

Where:

$$y(x_n,w) = \sigma\left(w' \emptyset(x_n)\right) = \frac{1}{1+\exp^{-w'\emptyset(x_n)}} \qquad (2\text{-}9)$$

In the study, multiple-kernel wavelet relevance vector machine classifier is expressed below.

The optimal offset b^* can be obtained by the following equation:

$$b^* = y_k - \sum_{i=1}^{n} \alpha_i^* y_i \left(\sum_{s=1}^{M} \mu_s^* k_s(x_k,x_i)\right) \qquad (2\text{-}10)$$

Then, the decision function obtained from the multiple-kernel RVM-based classifier can be expressed as followings:

$$f(x) = \text{sign}\left[\sum_{i=1}^{n} \alpha_i^* y_i \left(\sum_{s=1}^{M} \mu_s^* k_s(x_k,x_i) + b\right)\right] \qquad (2\text{-}11)$$

The wavelet function is described as followings:

$$K(x,x') = \sum_{i=1}^{l} p\left(\frac{x_i - x_i'}{\alpha}\right) \qquad (2\text{-}12)$$

In this reason, used for:

$$\psi(x) = \cos\left(\frac{2.5x}{2}\right)\exp\left(\frac{-x^2}{2}\right) \qquad (2\text{-}13)$$

This *Mercer kernel using Morlet mother wavelet*.

And kernel function is:

$$K(x,x') = \prod_{i=1}^{n} \cos\left(\frac{2.5(x_i - x_i')}{2\alpha}\right)\exp\left(\frac{-\|x_i - x_i'\|}{2\alpha^2}\right) \qquad (2\text{-}14)$$

Wavelet support vector machine and standard support vector machine have the same configuration basically, and the difference between them is kernel function.

Therefore kernel function of wavelet support vector machine is:

$$K(x,x') = \prod_{i=1}^{n} [\psi(\frac{x_i - x_i'}{\alpha})]^p$$

$$K(x,x') = \prod_{i=1}^{n} \psi(\frac{x_i - b_i}{\alpha}).\psi(\frac{x_i' - b_i}{\alpha})$$

$a \in z$, b_i is shift vector

Thus, the decision-making function for classification is:

$$f(x) = \text{sgn}\left[\sum_{i=1}^{n} a_i y_i \prod_{j=1}^{n} \psi(\frac{x^j - x_i^j}{\alpha}) + b\right] \qquad (2\text{-}15)$$

Therefore:

$$f(x,x') = \text{sgn}\left[\sum_{j=1}^{m} a_j y_j \left(\prod_{i=1}^{n} \cos\left(\frac{2.5(x_i - x_i^j)}{2\alpha}\right)\right)\exp\left(\frac{-\|x_i - x_i^j\|}{2\alpha^2}\right) + b)\right]$$

(2-16)

Where *m* is the number of training samples x_j is the j^{th} training sample, y_j is the training object of the j^{th} training sample, and ***x*** is the test sample.

III. Fault decision using Information fusion by feature level fusion(kernel fusion)

The information fusion based on wavelet-entropy is to make wavelet packet transformation to preparation data fusion and decompose it into different resolution space. Activity measure can acquire certain feature information of the multi-resolution analysis coefficient of the input image, and decide which image has more obvious feature information. The general activity measure is a certain function relative to detail component amplitude [6]. The definition is：

International Journal of Scientific Engineering and Technology (ISSN : 2277-1581)
Volume No.3 Issue No.3, pp : 310-315 1 March 2014

$$a_j^{\varepsilon}(m,n) = \sum_x p(m+m', n+n') * |D_j^{\varepsilon}(m+m', n+n')|^k$$
(2-17)

D_j^{ε} is the detail component coefficient matrix and $a_j^{\varepsilon}(m,n)$ is the activity measure of $D_j^{\varepsilon}(m,n)$, p is the mask of window area and it is used to linear filter D_j^{ε} , The activity measures said above are calculated by the components of detailed components decomposed with every level fail the impact of its corresponding proximate components. So the suggestion of entropy activity measure, taking the impact of both detail and proximate components on the activity measure into attention, achieves the objective of improving the effect of fusion [7].Suppose p is the window mask of *j* th level's detail component:

$$p_{i,j} = \begin{bmatrix} p_{1,1} & \cdots & p_{1,m} \\ \vdots & \ddots & \vdots \\ p_{n,1} & \cdots & p_{n,m} \end{bmatrix} \quad (2\text{-}18)$$

For every $p_{i,j}$ .Suppose in *j* level approximation coefficient matrix $p_{i,j}$ is

$$H_{i,j}^{l} = \begin{bmatrix} h_{1,1} & \cdots & h_{1,m} \\ \vdots & \ddots & \vdots \\ h_{n,1} & \cdots & h_{n,m} \end{bmatrix} \quad (2\text{-}19)$$

(Normalizing every point is: $H_{i,j} = \frac{h_{i,j}}{\sum h_{i,j}}$)

According to this, the formula to calculate the window entropy of $H_{i,j}^{l}$ is: $H_{i,j} = -\sum_{i=1}^{n} H_{i,j} log_2^{H_{i,j}}$ (2-20)

Then for the j[th] level detail component of input, the basic decision making module adopted with information fusion algorithm is:

$$\omega_{A,j}^{\varepsilon}(m,n) = \begin{cases} 0 & M_{j,AB}^{\varepsilon}(m,n) \le \sigma \text{ and } a_{j,A}^{\varepsilon}(m,n) \le a_{j,B}^{\varepsilon}(m,n) \\ 1 & M_{j,AB}^{\varepsilon}(m,n) \le \sigma \text{ and } a_{j,A}^{\varepsilon}(m,n) > a_{j,B}^{\varepsilon}(m,n) \\ \frac{1}{n} - \frac{1 - M_{j,AB}^{\varepsilon}(m,n)}{1-\sigma} & M_{j,AB}^{\varepsilon}(m,n) \ge \sigma \text{ and } a_{j,A}^{\varepsilon}(m,n) \le a_{j,B}^{\varepsilon}(m,n) \\ \frac{1}{n} + \frac{1 - M_{j,AB}^{\varepsilon}(m,n)}{1-\sigma} & M_{j,AB}^{\varepsilon}(m,n) \ge \sigma \text{ and } a_{j,A}^{\varepsilon}(m,n) \ge a_{j,B}^{\varepsilon}(m,n) \end{cases}$$

The superscript ε is the directions that detail component represents, $\omega_{A,j}^{\varepsilon}(m,n)$ is the decision factor of the fusion algorithm, $a_{j,s}^{\varepsilon}(m,n)$ is the entropy activity measure as described and σ is relativity threshold value. $M_{j,AB}^{\varepsilon}(m,n)$ is the relative co efficiency of the input images

$$M_{j,AB}^{\varepsilon}(m,n) = \frac{\sum\sum D_{j,A}^{\varepsilon}(m+m',n+n') D_{j,B}^{\varepsilon}(m+m',n+n')}{\sum\sum |D_{j,A}^{\varepsilon}(m+m',n+n')|^2 + |D_{j,B}^{\varepsilon}(m+m',n+n')|^2} \quad (2\text{-}21)$$

The wavelet coefficient after fusion could be showed as:

$$D_j^{\varepsilon}(m,n) = \sum_{i=1}^{i \ne j} \omega_{j,i}^{\varepsilon}(m,n) D_{j,i}^{\varepsilon}(m,n) \quad (2\text{-}22)$$

In formula: $\sum_{i=1}^{n} \omega_{j,i}^{\varepsilon}(m,n) = 1$ and $D_j^{\varepsilon}(m,n)$ is detail component coefficient on j level, in the direction of ε [7]. Morlet wavelet kernel not only has translation orthogonally, but also approximates an arbitrary function in the square integral space, such as the classification function *f* (***x***). Since the Morlet wavelet kernel has the nonlinear mapping ability, MWSVM has a good adaptive classification decision making ability. However, the actual classification problems are usually required to solve multiclass classification. Three approaches of creating MSVM by training and combining several classes SVM classifier, one-against-all, one-against-one, and DAGSVM [5]. The earliest used implementation for MSVM classification is probably the one-against-all method [5]. The MSVM using one-against-all strategy can be constructed by applying the following procedure [6]:

1- Construct *k* binary SVM classifiers where $f_i(x), i = 1, \dots, N$ separates training data of class *i* from the other training data:

$sgn[f_i(x)] = 1$, if instance *x* belongs to class *i*
$sgn[f_i(x)] = -1$, Otherwise.

2- Construct the k-class MSVM classifier by choosing the class corresponding to the maximal value of functions $f_i(x)$ The decision function is:

$$\delta(x) = \arg[\max\{f_1(x), \dots, f_N(x)\}]$$

Thus, the determined decision function is:

$$\delta(x) = \arg[\max\{\sum_{i,j} f_{ij}(x)\}]$$

Decision strategy:

$$d(x) = \arg\ \max\{v_1, \dots, v_N\}$$
$$v_i = \sum_{i=1}^{N} \delta_{ij} \quad i = 1,2,\dots,D$$

D: derivation tree

$$\delta_{ij} = \begin{cases} 1 & d_j = i \\ 0 & d_j \ne i \end{cases}$$

Where *d* (*x*) is the final decision function

v_i is the obtained votes of class *i* and d_j is the output of the *j*th MSVM trained by using the *j*th data source.

Therefore:

$$\delta_{ij} = \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_N,x_1) \\ \vdots & \ddots & \vdots \\ k(x_1,x_D) & \cdots & k(x_N,x_D) \end{bmatrix}^T$$

D: depth of tree N: Sensors

Therefore new method for fault diagnosis schema is in figure 2.

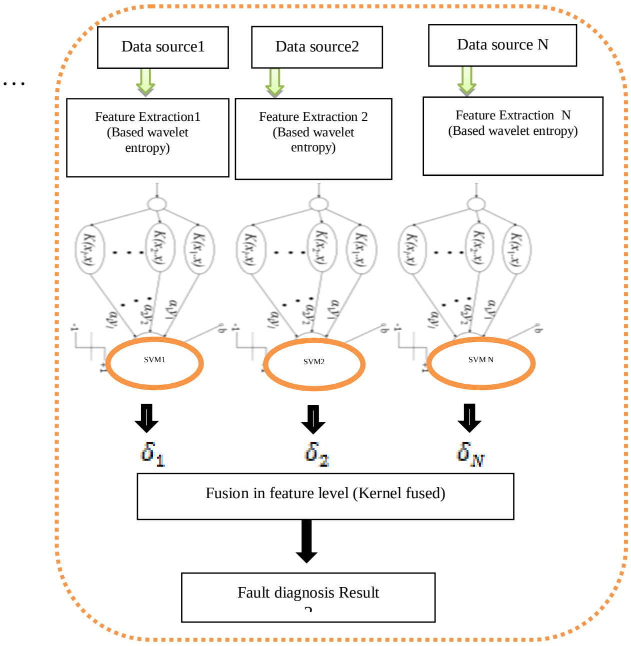

Fig 2: new method schema

This algorithm has best result because:

i. Time detection

Time detection is time derivation for extract data and decision. My method for feature extraction is Complex Wavelet Packet Entropy. This algorithm used Complex data for wavelet tree

and maximum coefficient on each node for signal optimization. Wavelet tree is fast method because coefficient of search algorithm in B-tree is 2^D with depth D.

$O(T_{CWP-E}) = 2^D$ D: depth of wavelet tree

ii. Time study:

In kernel method we have:

$$K(x,x')=\prod_{i=1}^{n}\cos\left(\frac{2.5(x_i-x_i')}{2a}\right)\exp\left(\frac{-\left\|x_i-x_i'\right\|}{2a^2}\right) \quad (3\text{-}9)$$

$$O(K) = n^3 e^{n}/2a^2$$

II. Results and Tables

We implementation this algorithm with input data whose get from teacher huang in mechanical school in WUT.

Step 1: input data

Data get from 2 sensors in gearbox machine

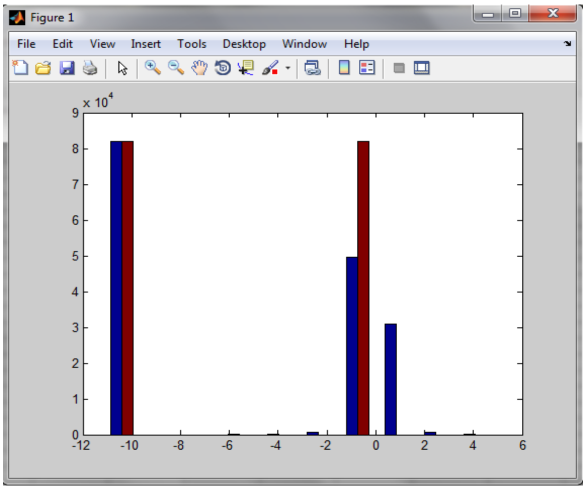

Figure 3: Data input from 2 sensors

Step 2: Feature Extraction based on wavelet Entropy

In this step extract data using wavelet packet (haar function with level 5 and selection data with max energy entropy

For sensor 1:

1. Feature extraction based on wavelet entropy

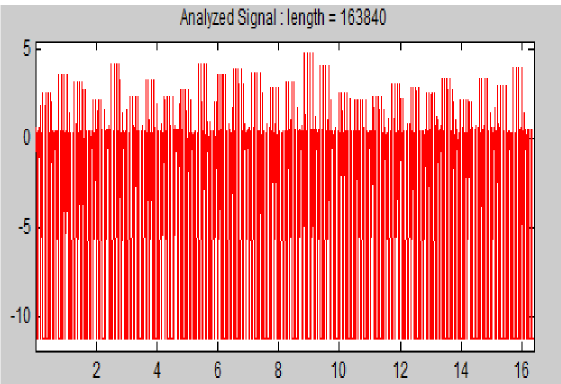

Figure 3-1: analysis input signal

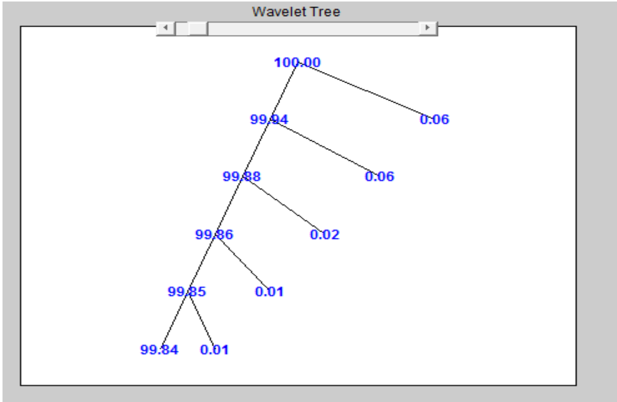

Figure 3-2: wavelet packet Tree

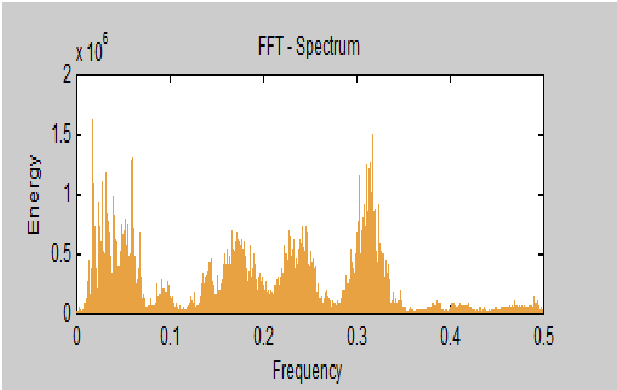

Figure 3-3: Max Energy Entropy

2. Classification using Kernel Method by wavelet

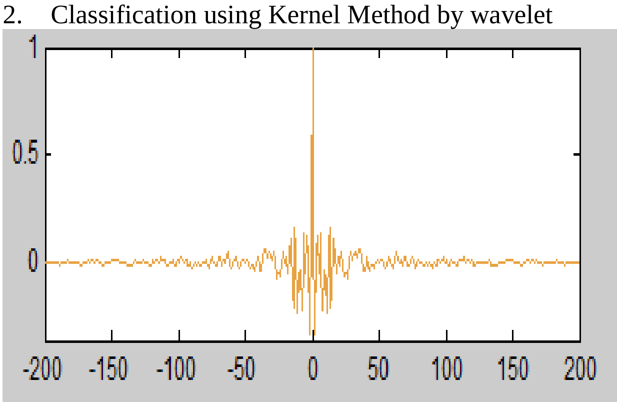

Figure 3-4: Kernel method

For sensor 2:

1. Feature extraction based on wavelet entropy

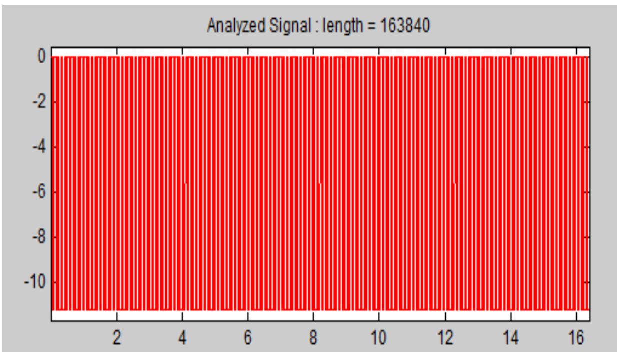

Figure 3-5: analysis input signal

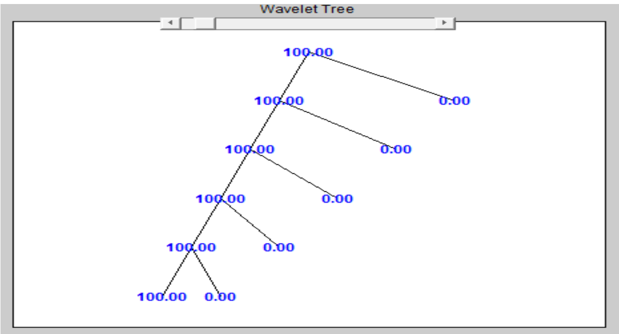

Figure 3-6: wavelet packet Tree

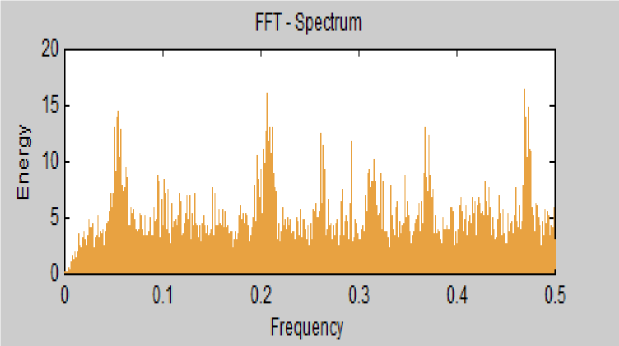

Figure 3-7: Max Energy Entropy

2. Classification using Kernel Method by wavelet

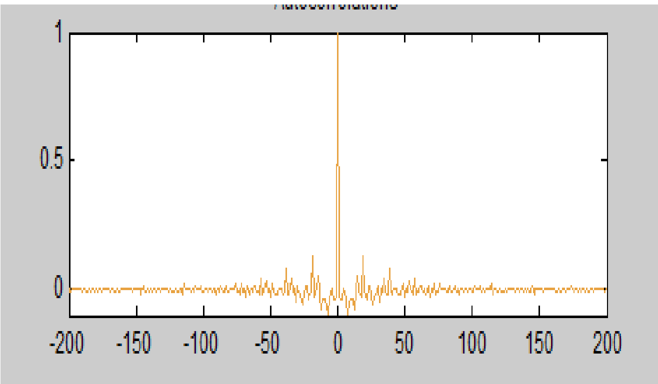

Figure 3-8: Kernel method

Step 3: Information Fusion (Kernel Fused)

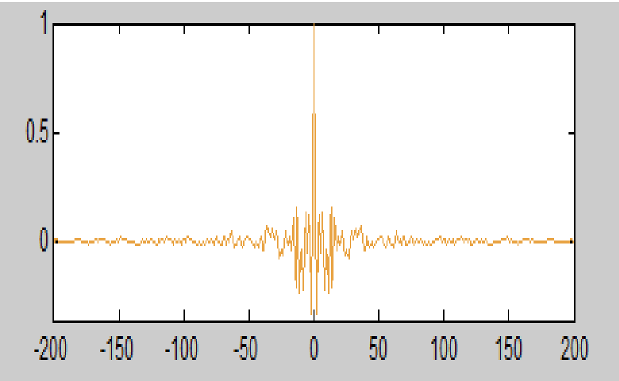

Figure 4: fused Kernel

Step 4: fault occupation

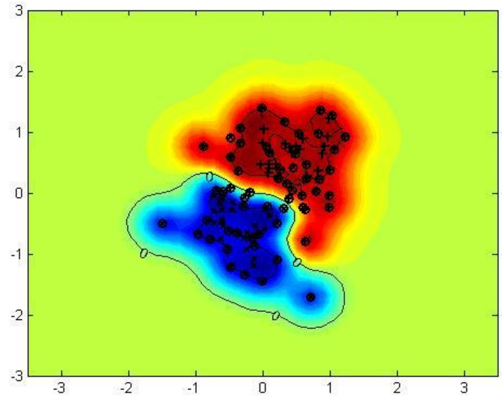

Figure 5: Kernel model Classification

Step 5: test data (confusion matrix)

Figure 6: test classification

IV. Conclusion

This algorithm offering for fault diagnosis with combine 3 method:

Once method is feature extraction using wavelet packet entropy. This algorithm used wavelet tree and maximum coefficient on each node for signal optimization. Wavelet tree is fast method because coefficient of search algorithm in B-tree is 2^D with depth D.

Second method is Multi classification. For this method suggested Kernel method with wavelet kernel (the kernel is MERCEL kernel using MORLET mother wavelet).for each input (sensor) make a SVM with kernel method. This method is best method for unsupervised learning model with minimum misclassification because in equation 2.16 we have $e_k = \frac{\alpha_k}{c}, \sum_{k=1}^{N} \alpha_k = 0$ therefore we must maximization normal vector of α_i otherwise we must maximized α_i ,this existence in this algorithm.

Third method is Information fusion. This new algorithm fused data on feature level. This method using maximization output of each SVM. The maximization model gets minimum time detection and time study in search model (Min O (n)).In kernel method we analysis kernel for fault size. Kernel method has 4 items for analysis in fault size:

a) Orientation of kernel
b) Number of peak
c) Size of a
d) Variance δ_l (deviation of the data with coordinate number *l)*

With select $\theta = 45^o$, Number of peak=3, δ_l=1, Kernel=Gaussian and a is optimization with kernel minimization.

Summary this algorithm is a best algorithm for fault diagnosis with multi input.

Acknowledgement

The authors thankful for Prof Hyang helpful for data and analyze data in mechanic school at Wuhan University of Technology.

References

i. *Shang, Xianghong Qian, Huanyan. A multi-kernel wavelet RVM algorithm for fault diagnosis of wireless sensor network. Journal of information and computational science, v9 n3, p711-717, March2012*

ii. *Zhang,Lei,Dong yanfei. Research on diagnosis of AC engine wear fault based on support vector machine and information fusion[j]. Journal of computers,v7,n9,p2292-2297-2012*

iii. *Vong,Chi-man; wong,pak-kin;lp,weng-fai. Case-based expert system using wavelet packet transform and kernel-based feature manipulation for engine ignition system diagnosis[j]. Engineering Application of Artificial Intelligence v24,n7, p2292-2297,2012*

iv. *liu,yu-ming;Ye,lu-bin;Zheng,Ping-you;Shi,Xiang-rong. Multiscale Classification and its Application to process monitoring[j]. Journal of Zhejiang University Science C,v11 ,n6, p425-534,june 2010*

v. *Yen, Gary G., Lin, Kuo-Chung, "Wavelet packet feature extraction for vibration monitoring", IEEE Transactions on Industrial Electronics, v 47, n 3, pp. 650-660, 2000.*

vi. *Mohammadreza vazifeh,Farzaneh abbasi,"Fault Diagnosis based on wavelet-Entropy and Information Fusion", Proceeding in 3th IEEE International conference on CSAE,guangzho,china,2013 DEC-in process*

vii. *Mohammad reza vazifeh,Pan Hao,Farzaneh abbasi; " Fault Diagnosis based on multi Kernel Classification and Information Fusion Decision"; journal of computer and technology,2013 NOV*